Supplementary information

Unbiased estimation of the OLS covariance matrix when the errors are clustered

Empirical Economics

Tom Boot, Gianmaria Niccodemi, and Tom Wansbeek (t.j.wansbeek@rug.nl)

Appendix A: Derivation of the unbiased variance estimators

Equicorrelated errors

In this section we consider the case where the errors are equicorrelated within clusters, so

$$\mathbf{\Sigma} = \sigma^2 \mathbf{I}_n + \tau^2 \mathbf{B}\mathbf{B}',$$

hence the design matrix for this case is

$$\mathbf{D} = (\text{vec }\mathbf{I}_n, \text{vec }\mathbf{B}\mathbf{B}').$$

Let

$$\begin{aligned} s &\equiv \text{tr}(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}} \\ \dot{s} &\equiv \text{tr}(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}} \\ \breve{s} &\equiv \text{tr}(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{X}}'\mathbf{\Delta}_n\tilde{\mathbf{X}}. \end{aligned}$$

Then

$$\begin{aligned} \mathbf{D}'\mathbf{D} &= \begin{pmatrix} \text{tr}\mathbf{I}_n & \text{tr}\mathbf{B}'\mathbf{B} \\ \text{tr}\mathbf{B}'\mathbf{B} & \text{tr}(\mathbf{B}'\mathbf{B})^2 \end{pmatrix} \\ &= \begin{pmatrix} n & n \\ n & \ddot{n} \end{pmatrix} \\ \mathbf{D}'(\mathbf{I}_n \otimes \mathbf{P})\mathbf{D} &= \begin{pmatrix} \text{tr}\mathbf{P} & \text{tr}\mathbf{B}'\mathbf{P}\mathbf{B} \\ \text{tr}\mathbf{B}'\mathbf{P}\mathbf{B} & \text{tr}\mathbf{B}'\mathbf{B}\mathbf{B}'\mathbf{P}\mathbf{B} \end{pmatrix} \\ &= \begin{pmatrix} k & s \\ s & \breve{s} \end{pmatrix} \\ \mathbf{D}'(\mathbf{P} \otimes \mathbf{P})\mathbf{D} &= \begin{pmatrix} \text{tr}\mathbf{P} & \text{tr}\mathbf{B}'\mathbf{P}\mathbf{B} \\ \text{tr}\mathbf{B}'\mathbf{P}\mathbf{B} & \text{tr}(\mathbf{B}'\mathbf{P}\mathbf{B})^2 \end{pmatrix} \\ &= \begin{pmatrix} k & s \\ s & \dot{s} \end{pmatrix}. \end{aligned}$$

So

$$\begin{aligned}\boldsymbol{\Psi} &\equiv \mathbf{D}'(\mathbf{M}\otimes\mathbf{M})\mathbf{D} \\ &= \begin{pmatrix} n-k & n-s \\ n-s & \ddot{n}-2\breve{s}+\dot{s}\end{pmatrix}.\end{aligned}$$

So for the current case (2) becomes

$$\begin{aligned}\hat{\mathbf{v}} &= \mathbf{R}'[\mathbf{D}'(\mathbf{M}\otimes\mathbf{M})\mathbf{D}]^{-1}\mathbf{D}'(\hat{\varepsilon}\otimes\hat{\varepsilon}) \\ &= (\mathbf{X}'\mathbf{X}\otimes\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}\otimes\mathbf{X})'\left(\text{vec }\mathbf{I}_n, \text{vec }\mathbf{B}\mathbf{B}'\right)\boldsymbol{\Psi}^{-1}\left(\text{vec }\mathbf{I}_n, \text{vec }\mathbf{B}\mathbf{B}')\right)'(\hat{\varepsilon}\otimes\hat{\varepsilon}) \\ &= (\mathbf{X}'\mathbf{X}\otimes\mathbf{X}'\mathbf{X})^{-1}\left(\text{vec }\mathbf{X}'\mathbf{X}, \text{vec }\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\boldsymbol{\Psi}^{-1}(\hat{\varepsilon}'\hat{\varepsilon}, \tilde{\hat{\varepsilon}}'\tilde{\hat{\varepsilon}})'.\end{aligned}$$

Cluster-specific parameters

We now let σ^2 and τ^2 vary over clusters and the parameter vector becomes

$$\boldsymbol{\lambda} = (\sigma_1^2, \ldots, \sigma_C^2, \tau_1^2, \ldots, \tau_C^2)'.$$

So now

$$\begin{aligned}\boldsymbol{\Sigma} &= \textstyle\sum_c(\sigma_c^2\mathbf{G}_c\mathbf{G}_c' + \tau_c^2\mathbf{b}_c\mathbf{b}_c') \\ \mathbf{D} &= \textstyle\sum_c(\mathbf{g}_c\mathbf{e}_c', \mathbf{h}_c\mathbf{e}_c'),\end{aligned}$$

with

$$\begin{aligned}\mathbf{g}_c &\equiv \text{vec}\mathbf{G}_c\mathbf{G}_c' \\ \mathbf{h}_c &\equiv \mathbf{b}_c\otimes\mathbf{b}_c,\end{aligned}$$

with properties

$$\begin{aligned}\mathbf{g}_c'\mathbf{g}_c &= n_c \\ \mathbf{h}_c'\mathbf{h}_c &= n_c^2 \\ \mathbf{g}_c'\mathbf{h}_c &= n_c,\end{aligned}$$

for $c = 1, \ldots, C$, while $\mathbf{g}_c'\mathbf{c}_d = \mathbf{h}_c'\mathbf{h}_d = \mathbf{c}_g'\mathbf{h}_d = 0$ for $d \neq c$, and

$$\begin{aligned}(\mathbf{X}\otimes\mathbf{X})'\mathbf{g}_c &= \text{vec}\mathbf{X}_c'\mathbf{X}_c \\ (\mathbf{X}\otimes\mathbf{X})'\mathbf{h}_c &= \tilde{\mathbf{x}}_c\otimes\tilde{\mathbf{x}}_c \\ (\hat{\varepsilon}\otimes\hat{\varepsilon})'\mathbf{g}_c &= \hat{\varepsilon}_c'\hat{\varepsilon}_c \\ (\hat{\varepsilon}\otimes\hat{\varepsilon})'\mathbf{h}_c &= \tilde{\hat{\varepsilon}}_c^2,\end{aligned}$$

with ε_c the residuals of cluster c and $\bar{\tilde{\varepsilon}}_c$ their sum over the observations in the cluster, this all for $c = 1, \ldots, C$. Further

$$
\begin{aligned}
\mathbf{g}_c'(\mathbf{I}_n \otimes \mathbf{P})\mathbf{g}_c &= (\mathrm{vec}\mathbf{G}_c\mathbf{G}_c')' \left(\mathbf{I}_n \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)(\mathrm{vec}\mathbf{G}_c\mathbf{G}_c') \\
&= \mathrm{tr}\left(\mathbf{G}_c\mathbf{G}_c'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}_c\mathbf{G}_c'\right) \\
&= \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_c'\mathbf{X}_c \\
&\equiv s_c \\
\mathbf{h}_c'(\mathbf{I}_n \otimes \mathbf{P})\mathbf{h}_c &= (\mathbf{b}_c \otimes \mathbf{b}_c)' \left(\mathbf{I}_n \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)(\mathbf{b}_c \otimes \mathbf{b}_c) \\
&= n_c\tilde{\mathbf{x}}_c'(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}_c \\
&\equiv n_c\tilde{s}_c \\
\mathbf{g}_c'(\mathbf{I}_n \otimes \mathbf{P})\mathbf{h}_c &= (\mathrm{vec}\mathbf{G}_c\mathbf{G}_c')' \left(\mathbf{I}_n \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)(\mathbf{b}_c \otimes \mathbf{b}_c) \\
&= \mathrm{tr}\left(\mathbf{G}_c\mathbf{G}_c'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{b}_c\mathbf{b}_c'\right) \\
&= \tilde{\mathbf{x}}_c'(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}_c \\
&= \tilde{s}_c,
\end{aligned}
$$

while it appears directly from the derivations that the terms across clusters are zero. This does not hold for the terms involving $\mathbf{P} \otimes \mathbf{P}$. There we have

$$
\begin{aligned}
\mathbf{g}_c'(\mathbf{P} \otimes \mathbf{P})\mathbf{g}_d &= (\mathrm{vec}\mathbf{G}_c\mathbf{G}_c')' \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)(\mathrm{vec}\mathbf{G}_d\mathbf{G}_d') \\
&= \mathrm{tr}\left(\mathbf{G}_c\mathbf{G}_c'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}_c\mathbf{G}_c'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \\
&= \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_c'\mathbf{X}_c(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_d'\mathbf{X}_d \\
&\equiv a_{cd} \\
\mathbf{h}_c'(\mathbf{P} \otimes \mathbf{P})\mathbf{h}_d &= (\mathbf{b}_c \otimes \mathbf{b}_c)' \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)(\mathbf{b}_d \otimes \mathbf{b}_d) \\
&= \left(\tilde{\mathbf{x}}_c'(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}_d\right)^2 \\
&\equiv q_{cd} \\
\mathbf{g}_c'(\mathbf{P} \otimes \mathbf{P})\mathbf{h}_d &= (\mathrm{vec}\mathbf{G}_c\mathbf{G}_c')' \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)(\mathbf{b}_d \otimes \mathbf{b}_d) \\
&= \mathrm{tr}\left(\mathbf{G}_c\mathbf{G}_c'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{b}_c\mathbf{b}_c\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)' \\
&= \tilde{\mathbf{x}}_d'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_c'\mathbf{X}_c(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}_d \\
&\equiv \ell_{cd}
\end{aligned}
$$

We let $\mathbf{\Delta}_s$ and $\mathbf{\Delta}_{\tilde{s}}$ be the diagonal matrices containing the s_c and $\tilde{s}_c$ and collect the a_{cd}, ℓ_{cd} and q_{cd} in the matrices $\mathbf{A}, \mathbf{L}$ and $\mathbf{Q}$, respectively. Then we obtain

$$\begin{aligned} \mathbf{D}'\mathbf{D} &= \begin{pmatrix} \mathbf{\Delta}_n & \mathbf{\Delta}_n \\ \mathbf{\Delta}_n & \mathbf{\Delta}_n^2 \end{pmatrix} \\ \mathbf{D}'(\mathbf{I}_n \otimes \mathbf{P})\mathbf{D} &= \begin{pmatrix} \mathbf{\Delta}_s & \mathbf{\Delta}_{\tilde{s}} \\ \mathbf{\Delta}_{\tilde{s}} & \mathbf{\Delta}_n \mathbf{\Delta}_{\tilde{s}} \end{pmatrix} \\ \mathbf{D}'(\mathbf{P} \otimes \mathbf{P})\mathbf{D} &= \begin{pmatrix} \mathbf{A} & \mathbf{L} \\ \mathbf{L}' & \mathbf{Q} \end{pmatrix}. \end{aligned}$$

So

$$\begin{aligned} \mathbf{\Phi} &= \mathbf{D}'(\mathbf{M} \otimes \mathbf{M})\mathbf{D} \\ &= \begin{pmatrix} \mathbf{\Delta}_n - 2\mathbf{\Delta}_s + \mathbf{A} & \mathbf{\Delta}_n - 2\mathbf{\Delta}_{\tilde{s}} + \mathbf{L} \\ \mathbf{\Delta}_n - 2\mathbf{\Delta}_{\tilde{s}} + \mathbf{L}' & \mathbf{\Delta}_n^2 - 2\mathbf{\Delta}_n \mathbf{\Delta}_{\tilde{s}} + \mathbf{Q} \end{pmatrix}. \end{aligned}$$

Combining the various elements, our unbiased estimator of the covariance matrix of the estimated regression coefficients is

$$\begin{aligned} \hat{\mathbf{v}} &= \mathbf{R}'[\mathbf{D}'(\mathbf{M} \otimes \mathbf{M})\mathbf{D}]^{-1}\mathbf{D}'(\hat{\varepsilon} \otimes \hat{\varepsilon}) \\ &= (\mathbf{X}'\mathbf{X} \otimes \mathbf{X}'\mathbf{X})^{-1}(\mathbf{X} \otimes \mathbf{X})' \textstyle\sum_c (\mathbf{g}_c \mathbf{e}_c', \mathbf{h}_c \mathbf{e}_c')\mathbf{\Phi}^{-1} \textstyle\sum_c (\mathbf{g}_c \mathbf{e}_c', \mathbf{h}_c \mathbf{e}_c')'(\hat{\varepsilon} \otimes \hat{\varepsilon}) \\ &= (\mathbf{X}'\mathbf{X} \otimes \mathbf{X}'\mathbf{X})^{-1} \textstyle\sum_c \Big((\text{vec}\mathbf{X}_c'\mathbf{X}_c)\mathbf{e}_c', (\tilde{\mathbf{x}}_c \otimes \tilde{\mathbf{x}}_c)\mathbf{e}_c' \Big) \mathbf{\Phi}^{-1} \textstyle\sum_c (\mathbf{e}_c \hat{\varepsilon}_c' \hat{\varepsilon}_c, \mathbf{e}_c \tilde{\hat{\varepsilon}}_c^2)'. \end{aligned}$$

Unrestricted error correlation within clusters

We now consider the case where the errors correlate freely within clusters, in a way that differs over clusters. The structure of $\mathbf{\Sigma}$ thus is

$$\begin{aligned} \mathbf{\Sigma} &= \text{diag}\ \mathbf{\Lambda}_c \\ &= \textstyle\sum_c \mathbf{G}_c \mathbf{\Lambda}_c \mathbf{G}_c'. \end{aligned}$$

This is a quite general structure, involving many parameters. It may even seem too generous in parameters but it has the merit to encompass all kinds of generalizations of the cluster-specific structure of Section 3.2 like factor structures. Since

$$\text{vec}\mathbf{\Sigma} = \textstyle\sum_c (\mathbf{G}_c \otimes \mathbf{G}_c)\text{vec}\mathbf{\Lambda}_c,$$

the design matrix now is, using the $\dot{\otimes}$ notation introduced at the end of Section 2,

$$\begin{aligned} \mathbf{D} &= \Big(\mathbf{G}_1 \otimes \mathbf{G}_1, \ldots, \mathbf{G}_C \otimes \mathbf{G}_C\Big) \\ &= \textstyle\sum_c \mathbf{e}_c' \dot{\otimes} \mathbf{G}_c \otimes \mathbf{G}_c. \end{aligned}$$

Then, with

$$\mathbf{P}_c \equiv \mathbf{X}_c(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_c',$$

we obtain

$$
\begin{aligned}
\mathbf{D}'\mathbf{D} &= \textstyle\sum_c \mathbf{e}_c\mathbf{e}_c' \dot{\otimes} \mathbf{I}_c \otimes \mathbf{I}_c \\
\mathbf{D}'(\mathbf{I}_n \otimes \mathbf{P})\mathbf{D} &= \left(\textstyle\sum_c \mathbf{e}_c \dot{\otimes} \mathbf{G}_c' \otimes \mathbf{G}_c'\right)(\mathbf{I}_n \otimes \mathbf{P})\left(\textstyle\sum_c \mathbf{e}_c' \dot{\otimes} \mathbf{G}_c \otimes \mathbf{G}_c\right) \\
&= \textstyle\sum \mathbf{e}_c\mathbf{e}_c' \dot{\otimes} \mathbf{I}_c \otimes \mathbf{P}_c \\
\mathbf{D}'(\mathbf{P} \otimes \mathbf{I}_n)\mathbf{D} &= \textstyle\sum \mathbf{e}_c\mathbf{e}_c' \dot{\otimes} \mathbf{P}_c \otimes \mathbf{I}_c.
\end{aligned}
$$

In the previous two cases we had a limited amount of parameters. But now we are faced with a possibly very large number of parameters, so we use (3) rather than (2).

Elaborating the expressions for $\mathbf{A}$ and $\mathbf{F}$ in (3) for the current case we get

$$
\begin{aligned}
\mathbf{A} &= \mathbf{D}'\mathbf{D} - \mathbf{D}'(\mathbf{I}_n \otimes \mathbf{P})\mathbf{D} - \mathbf{D}'(\mathbf{P} \otimes \mathbf{I}_n)\mathbf{D} \\
&= \textstyle\sum_c \mathbf{e}_c\mathbf{e}_c' \dot{\otimes} (\mathbf{I}_c \otimes \mathbf{I}_c - \mathbf{I}_c \otimes \mathbf{P}_c - \mathbf{P}_c \otimes \mathbf{I}_c) \\
&\equiv \textstyle\sum_c \mathbf{e}_c\mathbf{e}_c' \dot{\otimes} \mathbf{A}_c \\
\mathbf{F} &= \mathbf{D}'(\mathbf{X} \otimes \mathbf{X}) \\
&= \textstyle\sum_c \mathbf{e}_c \dot{\otimes} \mathbf{X}_c \otimes \mathbf{X}_c \\
&\equiv \textstyle\sum_c \mathbf{e}_c \dot{\otimes} \mathbf{F}_c,
\end{aligned}
$$

with $\mathbf{A}_c$ and $\mathbf{F}_c$ implicitly defined. Then

$$
\begin{aligned}
\mathbf{F}_c'\mathbf{A}_c &= \mathbf{X}_c' \otimes \mathbf{X}_c' - \mathbf{X}_c' \otimes \mathbf{X}_c'\mathbf{X}_c(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_c' - \mathbf{X}_c'\mathbf{X}_c(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_c' \otimes \mathbf{X}_c' \\
&= \left(\mathbf{I}_{k^2} - \mathbf{I}_k \otimes \mathbf{X}_c'\mathbf{X}_c(\mathbf{X}'\mathbf{X})^{-1} - \mathbf{X}_c'\mathbf{X}_c(\mathbf{X}'\mathbf{X})^{-1} \otimes \mathbf{I}_k\right)\mathbf{F}_c' \\
&\equiv \mathbf{S}_c\mathbf{F}_c',
\end{aligned}
$$

with $\mathbf{S}_c$ of order $k^2 \times k^2$ implicitly defined, so $\mathbf{F}_c'\mathbf{A}_c^{-1} = \mathbf{S}_c^{-1}\mathbf{F}_c'$ and

$$
\begin{aligned}
\mathbf{F}'\mathbf{A}^{-1}\mathbf{F} &= \textstyle\sum_c \mathbf{F}_c'\mathbf{A}_c^{-1}\mathbf{F}_c \\
&= \textstyle\sum_c \mathbf{S}_c^{-1}(\mathbf{X}_c'\mathbf{X}_c \otimes \mathbf{X}_c'\mathbf{X}_c).
\end{aligned}
$$

The final expression from (3) to be elaborated is

$$
\begin{aligned}
\mathbf{F}'\mathbf{A}^{-1}\mathbf{D}'(\hat{\varepsilon} \otimes \hat{\varepsilon}) &= \left(\textstyle\sum_c \mathbf{e}_c' \dot{\otimes} \mathbf{S}_c^{-1}\mathbf{F}_c'\right)\left(\textstyle\sum_c \mathbf{e}_c \dot{\otimes} \mathbf{G}_c' \otimes \mathbf{G}_c'\right)(\hat{\varepsilon} \otimes \hat{\varepsilon}) \\
&= \textstyle\sum_c \mathbf{S}_c^{-1}(\mathbf{X}_c'\hat{\varepsilon}_c \otimes \mathbf{X}_c'\hat{\varepsilon}_c).
\end{aligned}
$$

Then (3) becomes

$$
\hat{\mathbf{v}} = \left(\mathbf{X}'\mathbf{X} \otimes \mathbf{X}'\mathbf{X} + \textstyle\sum_c \mathbf{S}_c^{-1}(\mathbf{X}_c'\mathbf{X}_c \otimes \mathbf{X}_c'\mathbf{X}_c)\right)^{-1} \textstyle\sum_c \mathbf{S}_c^{-1}(\mathbf{X}_c'\hat{\varepsilon}_c \otimes \mathbf{X}_c'\hat{\varepsilon}_c).
$$

Appendix B: Degrees of freedom with random effects

In this appendix we elaborate the denominator of (19) and derive estimators for the parameters in $\hat{d}_\ell$. We start with the former. First,

$$\mathrm{tr}\mathbf{AM\Sigma MAM\Sigma M} = \sigma^4\mathrm{tr}\mathbf{AMAM} + 2\sigma^2\tau^2\mathrm{tr}\mathbf{B'MAMAMB} + \tau^4\mathrm{tr}(\mathbf{B'MAMB})^2. \tag{1}$$

The first term at the right-hand side was already elaborated in (18). As to the second term,

$$\mathbf{AMA} = \textstyle\sum_c \mathbf{G}_c\mathbf{A}_c^2\mathbf{G}_c' - \sum_{c,d}\mathbf{G}_c\mathbf{A}_c\mathbf{X}_c(\mathbf{X'X})^{-1}\mathbf{X}_d'\mathbf{A}_d\mathbf{G}_d'$$

so

$$\begin{aligned}\mathrm{tr}\mathbf{B'MAMAMB} &= \mathrm{tr}\textstyle\sum_c \mathbf{A}_c^2\mathbf{G}_c'\mathbf{MBB'MG}_c \\ &\quad -\mathrm{tr}\textstyle\sum_{c,d}(\mathbf{X'X})^{-1}\mathbf{X}_d'\mathbf{A}_d\mathbf{G}_d'\mathbf{MBB'MG}_c\mathbf{A}_c\mathbf{X}_c.\end{aligned}$$

From

$$\begin{aligned}\mathbf{G}_c'\mathbf{MB} &= \mathbf{i}_c\mathbf{e}_c' - \mathbf{X}_c(\mathbf{X'X})^{-1}\tilde{\mathbf{X}}' \\ &\equiv \mathbf{i}_c\mathbf{e}_c' - \mathbf{L}_c\end{aligned}$$

we obtain

$$\mathbf{G}_c'\mathbf{MBB'MG}_c = \mathbf{i}_c\mathbf{i}_c' - \mathbf{X}_c(\mathbf{X'X})^{-1}\tilde{\mathbf{x}}_c\iota_c' - \mathbf{i}_c\tilde{\mathbf{x}}_c'(\mathbf{X'X})^{-1}\mathbf{X}_c' + \mathbf{X}_c(\mathbf{X'X})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}(\mathbf{X'X})^{-1}\mathbf{X}_c'$$

and, letting $\boldsymbol{\mu}_c \equiv (\mathbf{X'X})^{-1}\mathbf{X}_c'\mathbf{A}_c\mathbf{i}_c$, we have $\mathrm{tr}\mathbf{B'MAMAMB} = T_1 + T_2$, with

$$\begin{aligned}T_1 &= \textstyle\sum_c \mathbf{i}_c'\mathbf{A}_c^2\mathbf{i}_c - 2\sum_c \mathbf{i}_c'\mathbf{A}_c^2\mathbf{X}_c(\mathbf{X'X})^{-1}\tilde{\mathbf{x}}_c + \mathrm{tr}\sum_c(\mathbf{X'X})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}(\mathbf{X'X})^{-1}\mathbf{X}_c'\mathbf{A}_c^2\mathbf{X}_c \\ T_2 &= \mathrm{tr}\textstyle\sum_{c,d}(\mathbf{X'X})^{-1}\mathbf{X}_d'\mathbf{A}_d(\mathbf{i}_d\mathbf{e}_d' - \mathbf{L}_d)(\mathbf{e}_c\mathbf{i}_c' - \mathbf{L}_c')e\mathbf{A}_c\mathbf{X}_c \\ &= \textstyle\sum_c \boldsymbol{\mu}_c'\mathbf{X'X}\boldsymbol{\mu}_c - 2\sum_c \tilde{\mathbf{x}}_c'(\mathbf{X'X})^{-1}\mathbf{X'AX}\boldsymbol{\mu}_c + \mathrm{tr}(\mathbf{X'X})^{-1}\mathbf{X'AX}(\mathbf{X'X})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}(\mathbf{X'X})^{-1}\mathbf{X'AX}.\end{aligned}$$

So far for the second term at the right-hand side of (27).

As to the third term, let $\lambda_c \equiv \mathbf{i}_c'\mathbf{A}_c\mathbf{i}_c$ and

$$\begin{aligned}\mathbf{B'MAMB} &= \textstyle\sum_c(\mathbf{B}' - \tilde{\mathbf{X}}(\mathbf{X'X})^{-1}\mathbf{X}')\mathbf{G}_c\mathbf{A}_c\mathbf{G}_c'(\mathbf{B} - \mathbf{X}(\mathbf{X'X})^{-1}\tilde{\mathbf{X}}') \\ &= \textstyle\sum_c\left(\lambda_c\mathbf{e}_c\mathbf{e}_c' - \tilde{\mathbf{X}}\boldsymbol{\mu}_c\mathbf{e}_c' - \mathbf{e}_c\boldsymbol{\mu}_c'\tilde{\mathbf{X}}'\right) + \tilde{\mathbf{X}}\mathbf{W}\tilde{\mathbf{X}}' \\ &\equiv \mathbf{S} + \tilde{\mathbf{X}}\mathbf{W}\tilde{\mathbf{X}}'.\end{aligned}$$

Then

$$\begin{aligned}\mathrm{tr}\mathbf{S}^2 &= \textstyle\sum_c\left(\lambda_c^2 - 4\mathbf{e}_c'\tilde{\mathbf{X}}\boldsymbol{\mu}_c + 2\boldsymbol{\mu}_c'\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\boldsymbol{\mu}_c\right) + 2\sum_{c,d}\tilde{\mathbf{x}}'\boldsymbol{\mu}_d\mathbf{e}_d'\tilde{\mathbf{x}}'\boldsymbol{\mu}_c \\ \mathrm{tr}\mathbf{S}\tilde{\mathbf{X}}\mathbf{W}\tilde{\mathbf{X}}' &= \textstyle\sum_c\left(\lambda_c\tilde{\mathbf{x}}_c'\mathbf{W}\tilde{\mathbf{x}}_c - 2\tilde{\mathbf{x}}_c'\mathbf{W}\tilde{\mathbf{X}}\boldsymbol{\mu}_c\right) \\ \mathrm{tr}(\tilde{\mathbf{X}}\mathbf{W}\tilde{\mathbf{X}}')^2 &= \mathrm{tr}\left(\mathbf{W}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^2.\end{aligned}$$

Combining these elements we obtain an expression for $\mathrm{tr}(\mathbf{B}'\mathbf{MAMB})^2$.

In the spirit of the "unbiased" theme of this paper, we estimate d_ℓ in (15) by using unbiased estimators for $\sigma^4, \sigma^2\tau^2$ and τ^4, which we will now derive. With the subscript to $\mathbf{m}_{ab}$ denoting an expression with a $\mathbf{M}$s and b $\mathbf{BB}'$s, there holds

$$\begin{aligned} \mathrm{E}(\hat{\varepsilon} * \hat{\varepsilon}) &= \mathbf{H}' \, \mathrm{E}(\hat{\varepsilon} \otimes \hat{\varepsilon}) \\ &= \mathbf{H}'(\mathbf{M} \otimes \mathbf{M})\mathrm{vec}(\sigma^2 \mathbf{I}_n + \tau^2 \mathbf{BB}') \\ &= \sigma^2 \mathbf{H}'\mathrm{vec}\mathbf{M} + \tau^2 \mathbf{H}'\mathrm{vec}\mathbf{MBB}'\mathbf{M} \\ &\equiv \sigma^2 \mathbf{m}_{10} + \tau^2 \mathbf{m}_{21}. \end{aligned}$$

We additionally have

$$\begin{aligned} \mathrm{E}(\hat{\varepsilon} * \mathbf{BB}'\hat{\varepsilon}) &= \mathbf{H}' \, \mathrm{E}(\hat{\varepsilon} \otimes \mathbf{BB}'\hat{\varepsilon}) \\ &= \mathbf{H}'(\mathbf{M} \otimes \mathbf{BB}'\mathbf{M})\mathrm{vec}(\sigma^2 \mathbf{I}_n + \tau^2 \mathbf{BB}') \\ &= \sigma^2 \mathbf{H}'\mathrm{vec}\mathbf{BB}'\mathbf{M} + \tau^2 \mathbf{H}'\mathrm{vec}\mathbf{BB}'\mathbf{MBB}'\mathbf{M} \\ &\equiv \sigma^2 \mathbf{m}_{11} + \tau^2 \mathbf{m}_{22} \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}(\mathbf{BB}'\hat{\varepsilon} * \mathbf{BB}'\hat{\varepsilon}) &= \mathbf{H}' \, \mathrm{E}(\mathbf{BB}'\hat{\varepsilon} \otimes \mathbf{BB}'\hat{\varepsilon}) \\ &= \mathbf{H}'(\mathbf{BB}'\mathbf{M} \otimes \mathbf{BB}'\mathbf{M})\mathrm{vec}(\sigma^2 \mathbf{I}_n + \tau^2 \mathbf{BB}') \\ &= \sigma^2 \mathbf{H}'\mathrm{vec}\mathbf{BB}'\mathbf{MBB}' + \tau^2 \mathbf{H}'\mathrm{vec}\mathbf{BB}'\mathbf{M}\,\mathbf{BB}'\mathbf{MBB}' \\ &\equiv \sigma^2 \mathbf{m}_{12} + \tau^2 \mathbf{m}_{23}. \end{aligned}$$

Then

$$\begin{aligned} \mathbf{i}_n' \, \mathrm{E}(\hat{\varepsilon} * \hat{\varepsilon} * \hat{\varepsilon} * \hat{\varepsilon}) &= 3\mathbf{i}_n' \left((\sigma^2 \mathbf{m}_{10} + \tau^2 \mathbf{m}_{21}) * (\sigma^2 \mathbf{m}_{10} + \tau^2 \mathbf{m}_{21}) \right) \\ \mathbf{i}_n' \, \mathrm{E}(\hat{\varepsilon} * \hat{\varepsilon} * \mathbf{BB}'\hat{\varepsilon} * \mathbf{BB}'\hat{\varepsilon}) &= \mathbf{i}_n' \left((\sigma^2 \mathbf{m}_{10} + \tau^2 \mathbf{m}_{21}) * (\sigma^2 \mathbf{m}_{12} + \tau^2 \mathbf{m}_{23}) \right. \\ &\quad \left. + 2(\sigma^2 \mathbf{m}_{11} + \tau^2 \mathbf{m}_{22}) * (\sigma^2 \mathbf{m}_{11} + \tau^2 \mathbf{m}_{22}) \right) \\ \mathbf{i}_n' \, \mathrm{E}(\mathbf{BB}'\hat{\varepsilon} * \mathbf{BB}'\hat{\varepsilon} * \mathbf{BB}'\hat{\varepsilon} * \mathbf{BB}'\hat{\varepsilon}) &= 3\mathbf{i}_n' \left((\sigma^2 \mathbf{m}_{12} + \tau^2 \mathbf{m}_{23}) * (\sigma^2 \mathbf{m}_{12} + \tau^2 \mathbf{m}_{23}) \right). \end{aligned}$$

Solving the sample counterpart of this system readily leads to unbiased estimators for the three parameters,

$$\begin{pmatrix} 3\mathbf{i}_n'(\mathbf{m}_{10} * \mathbf{m}_{10}) & 6\mathbf{i}_n'(\mathbf{m}_{10} * \mathbf{m}_{21}) & 3\mathbf{i}_n'(\mathbf{m}_{21} * \mathbf{m}_{21}) \\ x & y & z \\ 3\mathbf{i}_n'(\mathbf{m}_{12} * \mathbf{m}_{12}) & 6\mathbf{i}_n'(\mathbf{m}_{12} * \mathbf{m}_{23}) & 3\mathbf{i}_n'(\mathbf{m}_{23} * \mathbf{m}_{23}) \end{pmatrix} \begin{pmatrix} \widehat{\sigma^4} \\ \widehat{\sigma^2\tau^2} \\ \widehat{\tau^4} \end{pmatrix} = \begin{pmatrix} \sum_i \hat{\varepsilon}_i^4 \\ \sum_i \hat{\varepsilon}_i^2 \tilde{\varepsilon}_i^2 \\ \sum_i \tilde{\varepsilon}_i^4 \end{pmatrix},$$

with

$$\begin{aligned} x &\equiv \mathbf{i}_n'(\mathbf{m}_{10} * \mathbf{m}_{12} + 2\mathbf{m}_{11} * \mathbf{m}_{11}) \\ y &\equiv \mathbf{i}_n'(\mathbf{m}_{10} * \mathbf{m}_{23} + \mathbf{m}_{21} * \mathbf{m}_{12} + 4\mathbf{m}_{22} * \mathbf{m}_{11}) \\ z &\equiv \mathbf{i}_n'(\mathbf{m}_{21} * \mathbf{m}_{23} + 2\mathbf{m}_{22} * \mathbf{m}_{22}). \end{aligned}$$

Efficient computation can be based on $\mathbf{H}'\mathrm{vec}\mathbf{RS}' = (\mathbf{R} * \mathbf{S})\mathbf{i}_\ell$ for $\mathbf{R}$ and $\mathbf{S}$ of order $n \times \ell$.

Appendix C: Consistency in a simple setting

The simulations highlight that UV1 can offer a size correct test even with only a single treated cluster, while UV2 and UV3 require a somewhat larger number of treated clusters. In this section, we explore these findings from a theoretical perspective. We derive the conditions under which the three variance estimators are consistent in a simple model that is rich enough to explain the observed features from our simulations. We consider a setting with a single regressor, which is a treatment dummy that is equal to one in t_C out of C clusters. The design is balanced, so that each cluster contains n/C observations. With the definitions given in the paper, we then find

$$\mathbf{X}'\mathbf{X} = (n \cdot t_C)/C, \quad \tilde{\mathbf{X}}'\tilde{\mathbf{X}} = (n^2 \cdot t_C)/C^2. \tag{2}$$

We assume that the regression errors $\varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$. We take $\mathbf{\Sigma}$ as in Section 3.1, so that all variance estimators are unbiased. Without changing the proof, we can take $\mathbf{\Sigma}$ as in Section 3.2 and show consistency of UV2 and UV3. The relevant property of $\mathbf{\Sigma}$ is that its maximum eigenvalue satisfies $\lambda_{\max}(\mathbf{\Sigma}) \leq M \cdot n/C$. This condition holds in the specifications of Section 3.1 and Section 3.2. In this section M denotes a generic positive constant that can differ between appearances.

A sufficient condition for $\hat{v}/v \to_p 1$ is that $\text{var}(\hat{v})/v^2 \to 0$. Note that this implies that the degrees of freedom, defined as $2v^2/\text{var}(\hat{v})$, diverge. Using the expressions derived in Section 4, we have

$$\begin{aligned} \text{var}(\hat{v})/v^2 &= 2\text{tr}(\mathbf{AM\Sigma MAM\Sigma M})/v^2 \\ &\leq M \cdot \lambda_{\max}(\mathbf{\Sigma})^2(\mathbf{X}'\mathbf{X})^2\text{tr}(\mathbf{A}^2) \\ &\leq M \cdot (n/C)^4 t_C^2 \text{tr}(\mathbf{A}^2), \end{aligned} \tag{3}$$

where $\mathbf{A}$ is the block diagonal matrix with blocks $\mathbf{A}_c$ as given in Section 4 in the paper. We now proceed to derive explicit expressions for $\text{tr}(\mathbf{A}^2)$ for the variance estimators UV1-UV3.

UV1 From Section 4, we have that

$$\mathbf{A}_c = r_1\mathbf{I}_c + r_2\mathbf{i}_c\mathbf{i}_c', \quad (r_1, r_2) = \mathbf{f}_\ell'(\mathbf{X}'\mathbf{X} \otimes \mathbf{X}'\mathbf{X})^{-1}(\text{vec}\mathbf{X}'\mathbf{X}, \text{vec}\tilde{\mathbf{X}}'\tilde{\mathbf{X}})\mathbf{\Psi}^{-1} \tag{4}$$

The matrix $\mathbf{\Psi}$ defined in Section 3.1 depends on the following quantities,

$$\ddot{n} = \sum_c (n/C)^2 = n^2/C, \quad s = n/C, \quad \dot{s} = (n/C)^2, \quad \breve{s} = (n/C)^2.$$

Since $\mathbf{s}\Psi$ is a 2×2 matrix, its inverse is easily obtained as

$$\begin{aligned} \mathbf{\Psi}^{-1} &= \frac{1}{n}\begin{pmatrix} 1 - n^{-1} & 1 - C^{-1} \\ 1 - C^{-1} & (n/C)(1 - C^{-1}) \end{pmatrix} \\ &= \frac{1}{n}\frac{1}{(n/C - 1)(1 - C^{-1})}\begin{pmatrix} (n/C)(1 - C^{-1}) & -(1 - C^{-1}) \\ -(1 - C^{-1}) & 1 - n^{-1} \end{pmatrix}. \end{aligned}$$

Then, using (2),

$$(\text{vec}\mathbf{X}'\mathbf{X}, \text{vec}\tilde{\mathbf{X}}'\tilde{\mathbf{X}})\mathbf{\Psi}^{-1} = \frac{1}{C}\frac{1}{(n/C - 1)(1 - C^{-1})}(0, t_C(n/C - 1)) = (0, t_C/(C - 1)).$$

Using (4), we find that $r_1 = 0$ and $r_2 = C^2/(n^2 \cdot (C-1) \cdot t_C)$. Hence,

$$\mathrm{tr}(\mathbf{A}^2) = \sum_c \mathrm{tr}(\mathbf{A}_c^2) = C \cdot r_2^2 \cdot (n/C)^2 = \frac{C^2}{t_C^2 n^2} \frac{C}{(C-1)^2}.$$

Substituting this into (3), we find that

$$\mathrm{var}(\hat{v})/v^2 \leq M \cdot \left(\frac{n}{C}\right)^2 \frac{C}{(C-1)^2}.$$

We conclude that for UV1 to be consistent, we require that the number of clusters grows sufficiently fast to guarantee that $n^2/C^3 \to 0$. Importantly, consistency only depends on the number of clusters, and not on the number of treated clusters t_C.

UV2 Assume that the number of treated clusters $t_C > 2$. From Section 4, we have

$$\mathbf{A}_c = r_{1c}\mathbf{I}_c + r_{2c}\mathbf{i}_c\mathbf{i}_c', \quad (\mathbf{r}_1', \mathbf{r}_2') = \mathbf{f}_\ell'(\mathbf{X}'\mathbf{X} \otimes \mathbf{X}'\mathbf{X})^{-1} \sum_c \left((\mathrm{vec}\mathbf{X}_c'\mathbf{X}_c)\mathbf{e}_c', (\tilde{\mathbf{x}}_c \otimes \tilde{\mathbf{x}}_c)\mathbf{e}_c'\right) \mathbf{\Phi}^{-1}. \qquad (5)$$

Note that in the definition of $\mathbf{\Phi}$ given in Section 3.2 the elements $a_{cd}, \ell_{cd}, q_{cd}$ equal zero when one of the clusters is untreated. As a result, we have

$$\mathbf{\Phi} = \begin{pmatrix} \left(\frac{n}{C} - \frac{2}{t_C}\right)\mathbf{I}_{t_C} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}' & \mathbf{0} & \frac{n}{C}\left[\frac{t_C-2}{t_C}\mathbf{I}_{t_C} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}'\right] & \mathbf{0} \\ \mathbf{0} & \frac{n}{C}\mathbf{I}_{C-t_C} & \mathbf{0} & \frac{n}{C}\mathbf{I}_{C-t_C} \\ \frac{n}{C}\left[\frac{t_C-2}{t_C}\mathbf{I}_{t_c} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}'\right] & \mathbf{0} & \left(\frac{n}{C}\right)^2\left[\frac{t_C-2}{t_C}\mathbf{I}_{t_C} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}'\right] & \mathbf{0} \\ \mathbf{0} & \frac{n}{C}\mathbf{I}_{C-t_C} & \mathbf{0} & \left(\frac{n}{C}\right)^2\mathbf{I}_{C-t_C} \end{pmatrix}.$$

We can analytically invert this matrix by by rearranging it into a block diagonal matrix and first invert the diagonal blocks that are formed by

$$\mathbf{B}_1 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n}{C} \end{pmatrix} \begin{pmatrix} \left(\frac{n}{C} - \frac{2}{t_C}\right)\mathbf{I}_{t_C} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}' & \frac{t_C-2}{t_C}\mathbf{I}_{t_C} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}' \\ \frac{t_C-2}{t_C}\mathbf{I}_{t_c} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}' & \frac{t_C-2}{t_C}\mathbf{I}_{t_C} + \frac{1}{t_C^2}\mathbf{i}_{t_C}\mathbf{i}_{t_C}' \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{n}{C} \end{pmatrix}.$$

and

$$\mathbf{B}_2 = \begin{pmatrix} \frac{n}{C}\mathbf{I}_{C-t_C} & \frac{n}{C}\mathbf{I}_{C-t_C} \\ \frac{n}{C}\mathbf{I}_{C-t_C} & \left(\frac{n}{C}\right)^2\mathbf{I}_{C-t_C} \end{pmatrix}.$$

The inverse of the first diagonal block is simplified by that fact that the upper right, lower left and lower right blocks are identical. We obtain

$$\mathbf{B}_1^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{C}{n} \end{pmatrix} \begin{pmatrix} \left(\frac{n}{C} - 1\right)^{-1}\mathbf{I}_{t_C} & -\left(\frac{n}{C} - 1\right)^{-1}\mathbf{I}_{t_C} \\ -\left(\frac{n}{C} - 1\right)^{-1}\mathbf{I}_{t_C} & \left(\left(\frac{n}{C} - 1\right)^{-1} + \frac{t_C}{t_C-2}\right)\mathbf{I}_{t_C} - \frac{1}{(t_C-2)(t_C-1)}\mathbf{i}_{t_C}\mathbf{i}_{t_C}' \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{C}{n} \end{pmatrix},$$

The inverse of the second block is

$$\mathbf{B}_2^{-1} = \begin{pmatrix} \frac{n}{C}\mathbf{I}_{C-t_C} & \frac{n}{C}\mathbf{I}_{C-t_C} \\ \frac{n}{C}\mathbf{I}_{C-t_C} & \left(\frac{n}{C}\right)^2\mathbf{I}_{C-t_C} \end{pmatrix}^{-1} = \frac{1}{\frac{n}{C} - 1} \begin{pmatrix} \mathbf{I}_{C-t_C} & -\frac{C}{n}\mathbf{I}_{C-t_C} \\ -\frac{C}{n}\mathbf{I}_{C-t_C} & \frac{C}{n}\mathbf{I}_{C-t_C} \end{pmatrix}$$

Rearranging back into the original form of $\mathbf{\Phi}$, we obtain

$$\mathbf{\Phi}^{-1} = \left(\frac{n}{C} - 1\right)^{-1} \begin{pmatrix} \mathbf{I}_{t_C} & \mathbf{0} & -\frac{C}{n}\mathbf{I}_{t_C} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{C-t_C} & \mathbf{0} & -\frac{C}{n}\mathbf{I}_{C-t_C} \\ -\frac{C}{n}\mathbf{I}_{t_C} & \mathbf{0} & \left(\frac{n}{C}-1\right)\left(\frac{C}{n}\right)^2\left[\left(\left(\frac{n}{C}-1\right)^{-1} + \frac{t_C}{t_C-2}\right)\mathbf{I}_{t_C} - \frac{1}{(t_C-2)(t_C-1)}\mathbf{i}_{t_C}\mathbf{i}'_{t_C}\right] & \mathbf{0} \\ \mathbf{0} & -\frac{C}{n}\mathbf{I}_{C-t_C} & \mathbf{0} & \frac{C}{n}\mathbf{I}_{C-t_C} \end{pmatrix}.$$

Using that for the treated clusters $\mathbf{X}_c'\mathbf{X}_c = n/C$ and $\tilde{\mathbf{x}}_c \otimes \tilde{\mathbf{x}}_c = (n/C)^2$, we get that

$$\begin{aligned} \sum_c \left((\text{vec}\mathbf{X}_c'\mathbf{X}_c)\mathbf{e}_c', (\tilde{\mathbf{x}}_c \otimes \tilde{\mathbf{x}}_c)\mathbf{e}_c'\right)\mathbf{\Phi}^{-1} &= \left(\mathbf{0}_C', \left(\frac{t_C}{t_C-2} - \frac{t_C}{(t_C-2)(t_C-1)}\right)\mathbf{i}'_{t_C}, \mathbf{0}'_{C-t_C}\right) \\ &= \left(\mathbf{0}_C', \frac{t_C}{t_C-1}\mathbf{i}'_{t_C}, \mathbf{0}'_{C-t_C}\right). \end{aligned}$$

In (5) we now have $\mathbf{r}_1 = \mathbf{0}$ and $\mathbf{r}_{2c} = \frac{1}{t_C^2}\left(\frac{C}{n}\right)^2\frac{t_C}{t_C-1}$ if cluster c is treated and zero otherwise. Then,

$$\text{tr}(\mathbf{A}^2) = t_C\left(\frac{n}{C}\right)^2 \cdot \frac{1}{t_C^4}\left(\frac{C}{n}\right)^4\frac{t_C^2}{(t_C-1)^2}.$$

Substituting this into (3) we conclude that

$$\text{var}(\hat{v})/v^2 \leq M \cdot \left(\frac{n}{C}\right)^2\frac{t_C}{(t_C-1)^2}.$$

We conclude that UV2 is consistent for v when $n^2/(C^2 \cdot t_C) \to 0$. A necessary condition for this to happen is that the number of treated clusters $t_C \to \infty$. This stands in marked contrast with the finding for UV1 above.

UV3 Assume that $t_C > 2$.

$$\mathbf{A}_c = \mathbf{X}_c\mathbf{Q}_c\mathbf{X}_c', \quad (\text{vec}\mathbf{Q}_c)' = \mathbf{f}_\ell'\left(\mathbf{X}'\mathbf{X} \otimes \mathbf{X}'\mathbf{X} + \textstyle\sum_c \mathbf{S}_c^{-1}(\mathbf{X}_c'\mathbf{X}_c \otimes \mathbf{X}_c'\mathbf{X}_c)\right)^{-1}\mathbf{S}_c^{-1}. \tag{6}$$

If cluster c is treated, we have $S_c = 1 - 2/t_C$, while if cluster c is not treated, we have $S_c = 1$. Then,

$$\mathbf{X}'\mathbf{X} \otimes \mathbf{X}'\mathbf{X} + \textstyle\sum_c \mathbf{S}_c^{-1}(\mathbf{X}_c'\mathbf{X}_c \otimes \mathbf{X}_c'\mathbf{X}_c) = (n \cdot t_C)^2/C^2 + t_C^2/(t_C-2)(n/C)^2 = \left(\frac{n \cdot t_C}{C}\right)^2 \cdot \frac{t_C-1}{t_C-2}.$$

Then with A_c from (6), we find

$$\text{tr}(\mathbf{A}^2) = \sum_c \text{tr}(\mathbf{A}_c^2) = t_C \cdot \frac{n^2}{C^2}\frac{C^4}{n^4 \cdot t_C^4}\frac{(t_C-2)^2}{(t_C-1)^2}\frac{t_C^2}{(t_C-2)^2} = \frac{C^2}{n^2 t_C(t_C-1)^2}.$$

Substituting this into (3), we conclude that

$$\text{var}(\hat{v})/v^2 \leq M \cdot \left(\frac{n}{C}\right)^2\frac{t_C}{(t_C-1)^2}.$$

For consistency of UV3 we therefore require that $n^2/(C^2 \cdot t_C) \to 0$. This is the same condition as for UV2, so that again we require the number of treated clusters $t_C \to \infty$.